DOCUMENT RESUME

ED 054 220 TM 000 796

AUTHOR Hartnett, Rodney T.
TITLE Evaluation, Accountability, and a Consideration of Some of the Problems of Assessing College Impacts.
INSTITUTION Educational Testing Service, Princeton, N.J.
REPORT NO RM-71-7
PUB DATE Jun 71
NOTE 27p.

EDRS PRICE EDRS Price MF-$0.65 HC-$3.29
DESCRIPTORS Behavioral Objectives, *College Planning, College Role, *Educational Accountability, *Educational Improvement, *Educational Objectives, Educational Quality, *Evaluation, Evaluation Methods, Financial Problems, Higher Education, Information Systems, Public Opinion, Relevance (Education), Research Problems

ABSTRACT

The terms educational accountability and educational evaluation are clarified, and problems involved in assessing college effectiveness and developing objective criteria for accountability are discussed. (MS)

EDO 54220

RESEARCH MEMORANDUM

RM-71-7

EVALUATION, ACCOUNTABILITY, AND A CONSIDERATION OF SOME OF THE PROBLEMS OF ASSESSING COLLEGE IMPACTS

Rodney T. Hartnett

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

This Memorandum is for interoffice use. It is not to be cited as a published report without the specific permission of the author.

TM 000 796

Educational Testing Service
Princeton, New Jersey
June 1971

Evaluation, Accountability, and a Consideration of Some of the Problems of Assessing College Impacts

Rodney T. Hartnett

Educational Testing Service

Almost all reasonable observers of American higher education agree that the time has arrived--indeed has been with us right along, though too few have been aware of it--for higher education to take a close, careful, and critical examination of itself. While it is true that there has always been the need for institutions to conduct ongoing programs of self-evaluation, the external pressures (that is, pressures from public officials, potential donors to the institution, tax-payers, etc.) for colleges and universities to take stock of themselves is greater now than perhaps ever before in the history of American higher education.

There are many reasons for the increased demand for institutional self-scrutiny, of course. One of the most important, especially in the public sector, is the fantastic increase in consolidated systems of higher education in the past decade. It would appear that the crucial years were 1960 and 1961, when many states began to realize that voluntary planning and coordinating efforts were not going to be sufficient to meet the challenges of the 1960's.[1] At that time several states either enacted legislation creating mandatory coordinating and planning agencies, or strengthened the power of existing ones. The trend was thus set in motion, and the implications for statewide evaluation and systematic accounting procedures were clear. Statewide planning, if it were to be at all superior to the nearly

---

[1]Palola, Ernest G., Lehmann, Timothy, and Blischke, William R., Higher Education by Design: The Sociology of Planning. Center for Research & Development in Higher Education, University of California, Berkeley, 1970.

autonomous development of institutions that preceded it, had to be based on more than pure fancy. Institutions were now expected to justify their requests for money, approval for new programs and the like, on facts about their institutions and its operations. Thus, even though "institutional research" had been around for a long time, it was really during the early 60's that very many colleges and universities began to take it seriously. According to a survey conducted by Francis Rourke and Glenn Brooks, there were only 10 institutions of higher education in the country boasting formal offices of institutional research prior to 1955, but by 1964 the number had swelled to 115.[2]

Closely related to the growth of multi-instituional coordinating agencies has been the increasing financial problems confronting higher education, a fiscal shortage of growing urgency in the past five years which has recently reached crisis proportions. According to a recent report by the Carnegie Commission on Higher Education, "higher education has come upon hard times. The trouble is serious enough to be called a depression."[3] The same study goes on to predict that if the current trend continues, almost all higher educational institutions eventually will be in financial difficulty. Support for such a position is provided by a report from the American Association for Higher Education, which claims that not only is support for higher education descending rapidly, but also that there is no indication of a let-up in the money squeeze for the next five years.[4]

---

[2]Rourke, Francis, E., and Brooks, Glenn E. The Managerial Revolution in Higher Education. Baltimore: The John Hopkins Press, 1966.

[3]Cheit, Earl F. The New Depression in Higher Education. McGraw-Hill, 1971.

[4]Blackburn, Robert T. Changes in Faculty Life Styles, Research Report Number 1, American Association for Higher Education, 1970.

The Credibility Gap and Demands for Educational "Accounting"

The reasons offered for the financial crisis in higher education are numerous and often interrelated, including the Vietnam war, a national rearrangement of priorities (with greater attention going to poverty, racism, and ecological problems) increased enrollments, rising costs, and an overall steady decline of the American economy. Undoubtedly, however, one of the major causes of the current income shortage in higher education is what might be referred to as "the credibility gap," a growing feeling of mistrust on the part of higher education's relevant publics (be they alumni, parents of school-age children, or whatever) about what higher education is doing or "producing." Such uneasy feelings have been nurtured, of course, by the rash of campus disturbances during the past few years, disturbances which have led to adverse reactions affecting both private and legislative support. It would probably be a mistake, however, to lay a disproportionate share of the blame for the "credibility gap" at the foot of the campus protestors.[5] While they may have provided the observable stimulus for increased expressions of mistrust, it is probably safe to say that higher educational institutions have long been viewed with suspicion by many who have helped support them. Such misgivings are tolerable during periods when the economy is on the upswing. But during a questionable economy or a clear-cut recession, it is understandable that money finds its way to those who can demonstrate that the money has been well spent. While better times would have been characterized by a sort of suspicious laissez faire attitude toward higher education, we now see a

---

[5]Or too much credit, either. It is ironic to note that the mistrust for higher education arising from campus disruptions is to some extent a sign of the success of such demonstrations, for the purpose of many activist students is to highlight the lack of relevance and worthlessness of higher education generally.

demand for evidence that the large sum of money being spent on American higher education is being judiciously allocated. Concern about the costs of new educational programs, renewed interest in the costs of old programs, questions about the need for annual faculty salary increases, the legitimacy of the practice of tenure--all these and more are being critically reappraised. At all levels and in various ways higher educational institutions are being called upon to "account" for their programs and actions, just as other institutions or agencies are expected to justify their operations. College administrators, who have been allowed to luxuriate in the secrecy of their tasks, are now being pressured into a stance of openness. All who make claims for their "products" are asked to provide evidence to support their claim, and although there are numerous other reasons for institutions to carefully and systematically study themselves, it is quite clear that financial stress is the most powerful persuader.

It is also clear that the "institutional research" which has been carried out continually on many campuses and the kind of educational accounting that is being demanded of higher education now are not one and the same. In a broad sense, of course, they are both forms of educational evaluation, a practice which has been around for many years, but evaluation and "accountability" are not the same either, even though, again, the overlap between the concepts is substantial. Nor is accountability synonymous with "management information systems," "cost-benefit analysis," or "Program Planning and Budgeting Systems," though all of these are interrelated. Consequently, it is imperative that the distinctions between and among these various concepts be clarified.

An Attempt at Some Conceptual Clarifications

Evaluation in Higher Education

Evaluation in higher education has traditionally been concerned with how well or to what degree specifically defined objectives of a program (a curriculum, a set of operating principles, or whatever) were attained. In a small percentage of cases the essential ingredients of such an undertaking have been very much like those employed by a scientist (social or other): 1) behaviorally defined objectives, 2) the random assignment of subjects (usually educational experiences), 3) clearly differentiated treatments (such as different teaching techniques or other forms of curricular innovation), and 4) criterion measures chosen or developed on the basis of the behavioral objectives. Most programs in higher education, however, have not lent themselves to this experimental model. Obviously, it is quite insensitive to most of the "real world" problems confronted in higher education. As one evaluator has remarked: "What does one do when not all the relevant objectives are manifested in directly observable specific individual behavior? What does one do about deliberately trying to measure effects that are not objectives of the program? What does one do when random assignment of subjects to treatments cannot be accomplished? What does one do when he lacks clearly differentiated treatments?"[6] Because of concerns such as these, most educational evaluation has been based on a model which is both more comprehensive and more flexible. The two outstanding features of this model have been, first, a concern with the question "what are the consequences

---

[6]Pace, C. Robert. An Evaluation of Higher Education: Plans and Perspectives, CSE Report No. 51, Center for the Study of Evaluation, UCLA Graduate School of Education, January, 1969, page 2 (mimeo).

of higher education?" (rather than the objectives), and second, a style of inquiry which is more exploratory in nature (as opposed to the experimental orientation of the classical model). The concern with the consequences of higher education stems from recognition that certain outcomes of higher education are often unintended (or at least not specifically stated) but still potentially important, and to ignore them simply because they were not acknowledged at the outset would be to neglect important and illuminating information. The preference for a style of inquiry which is exploratory in nature emerges from an awareness that higher educational institutions are not scientific laboratories in which the various elements of the enterprise can be carefully controlled or manipulated to please the evaluator. Many institutions are continually changing their programs, toying with new approaches, and attempting to engender free environments. The exploratory style is typified by the comment: "The spirit of the evaluator should be adventurous. If only that which could be controlled or focused were evaluated, then a great many important educational and social developments would never be evaluated--at least not by 'evaluators;' that would be a pity."[7]

Educational Accounting

"Accountability" is the new "in" word in American education. The concept of educational accountability has been the subject of numerous symposia and special issues of educational journals, and certain forms of educational accountability have been brought to the attention of the American public through popular accounts in the newspaper and other news media. It is a very sensitive concept, one

---

[7] Ibid., p. 3.

which has been the center of much controversy at the elementary and secondary school levels.

In many ways, educational accountability and educational evaluation are essentially the same. Accountability, like evaluation, is aimed at learning about the effect of educational institutions. Like evaluation, accountability is concerned with the effect of certain educational "treatments" (school experiences) on the students, after relevant characteristics of the students at the time the students entered college are "controlled." The question "Are our institutions living up to their claims?" is of primary concern to both evaluators and accountability experts.

The differences between evaluation and accountability are less obvious, but very important. First of all, evaluation is concerned primarily with educational effectiveness (the degree to which it succeeds in doing whatever it is trying to do), whereas accountability experts are concerned with effectiveness and efficiency (its capacity to achieve results with a given expenditure of resources), and very often they are more interested in the latter. Thus, while the evaluator's task is an extremely difficult one (some of the difficulties will be discussed in the next section of this paper), the educational accountant's role is even more complex, for he not only attempts to determine what the institution has done, but also how much it has cost to do it, and ultimately, whether it was worth the cost.

Of course, as Rourke and Brooks point out, efficiency and effectiveness are closely related, for how well an institution achieves its goals may depend largely on how well it has used its usually limited resources.[8] But the two

---

[8]Rourke and Brooks, op.cit.

are often at odds, as demonstrated by the rather frequent clash between the college financial officer, who often tends to be oriented toward a criterion of efficiency, and the faculty member who complains about the restraints being placed on his strivings for educational effectiveness.

A second difference between evaluation and accountability has to do with the stimulus for the study and who participates in the inquiry. Institutional evaluation has traditionally been an activity carried out as an ongoing function within the institution by members of administrative and faculty groups. The entire process of self-study has been viewed as one which would enable members of the staff to gain more insights into their own strengths and weaknesses and thereby improve the educational, research and service programs of the institution. It is viewed, in other words, as having positive ends. Accountability, on the other hand, has brought with it the notion of external judgment. Judging, at least, from the reactions of many elementary and secondary school teachers, there is the clear indication that "accountability" is regarded as a vindictive rather than an affirmative process. Someone not in the school itself is passing judgment on the quality of the performance of those who work there. The terminology often found in articles and papers making a case for accountability, include such phrases as, "the professional educators who operate them (the schools) must be held responsible," "the taxpayers are entitled to know what they are getting," etc. As one teacher has remarked: "If we say that someone is accountable we usually mean that 'he must suffer the consequences of his actions' We hardly ever mean the more positive 'he will profit from the consequences of his actions.'"[9]

---

[9]McGhan, Barry R. Accountability as a Negative Reinforcer, American Teacher, 1970, Vol. 55, No. 3, November, p. 13.

Though there are other differences between evaluation and accountability (e.g., educational evaluators are often psychologists or educational researchers, whereas educational accountants are more often economists or from backgrounds in business and finance), the differences between effectiveness and efficiency as the focus of the research and between the perceptions (accurate or not) of evaluation as a positive form of self-study and accountability as a retributive form of judgment from some external body, seem to be the major distinguishing characteristics.

Educational accountability can and does take many forms. At the higher education level, two forms seem to be most likely to gain support. The first is for higher educational institutions (or systems) to move toward improved, output-oriented management methods, always with an eye toward efficiency. In many institutions, this has been the primary function of their Offices of Institutional Research for some years. The institutions perform their own self-study (as in evaluation), based on improved output-oriented management methods such as program budgeting (as opposed to straight line-item budgeting), systems analysis, standard forms for gathering basic institutional data (and routine computer programs to yield reports), etc. The institutions then make their own periodic reports to their relevant publics, e.g., their alumni or donors in the case of private institutions, the board of regents, statewide coordinating body, or whatever, in the case of public institutions.

The second form of accountability which would seem to be viable in higher educational institutions is what Stephen Barro calls "institutionalization of external evaluations or educational audits."[10] In this accountability system, assessments of efficiency and effectiveness would be made by some agency external to the

---

[10]Barro, Stephen M. An Approach to Developing Accountability Measures for the Public Schools, Phi Delta Kappan, 1970, XII, No. 4, December, p. 197.

institution, such as by a statewide office of higher education. In this case, the institution's performance would be judged by direct comparison to others with the same financial base. All data used for such comparisons would have to be objective and comparable among all institutions, such data being gathered by the central agency by means of standard reporting routines and kept in a central data file for purposes of regular inter-institutional comparisons.

A third form of accountability which might conceivably gain support among those passing judgment on the quality of an institution's activities is a performance incentive system for faculty members. Under this plan, salary increases, promotion, or other devices may be used as rewards for demonstrated quality performance by the faculty.[11] Such an approach would bring the accountability notion right down to specific members of the faculty, whereas it is usually thought of as pertaining to the institutional or possibly departmental level. Yet, the current overabundance of PhD's and scarcity of vacancies at the college level, combined with the growing insistence among students for ratings of their teachers, make it more likely that "accountability" at the individual teacher level may be forthcoming.

There are other forms of educational accountability, but their appropriateness for higher education is questionable. Performance contracting (in which contracts are made with external agencies--usually private firms--to conduct specified instructional activities presumably leading to agreed-upon, measurable results, such as a gain score of so many points on a standardized reading test), alternative

---

[11] Though some higher educational institutions have occasionally granted cash awards to faculty members voted as outstanding teachers by the students, such reinforcement is usually available to so few that it can hardly be regarded as a bona fide performance incentive system as meant here.

educational systems (also referred to as the "voucher system"), etc.--these seem to be less suited for higher education, mainly because they are geared to an educational level at which there is rather wide agreement or consensus about the specific developmental skills (e.g., reading, writing) expected of its students.

Management Information Systems

A central feature of accountability systems in higher education--especially the external evaluation by a central agency--is the management information system (MIS). The MIS is a system of information collection, storage, collating and distribution which makes it possible to routinely monitor certain aspects of an institution's operations. At the heart of the MIS is a central pool of data, consisting of compatible pieces of information. Such a system makes inter-institutional comparisons possible and meaningful, for the interpretations can be based on common data elements. One of the problems of making inter-institutional comparisons in the past has been that the information available has not been compatible. A full-time-equivalent student at one institution, for example, has not necessarily been the same as a full-time-equivalent student in another institution. And so on. Management information systems, in and of themselves, do not represent another form of educational accounting or evaluation. They are an indispensable tool, however, for the conduct of any form of inter-institutional comparisons.

A good example of an MIS for higher education is the one developed by the Systems Research Group of Toronto. Known by the acronym CAMPUS (for Comprehensive Analytical Methods for Planning University Systems), this MIS is designed to help colleges and universities "gain the maximum educational advantage from the resources which are put at their disposal."[12] CAMPUS focuses on basic operational data which

---

[12]The Development and Implementation of CAMPUS: A Computer-Based Planning and Budgeting Information System for Universities and Colleges. Systems Research Group, Toronto, Ontario, Canada (August, 1970, p. 2).

are already available, in some form, on most campuses. By concentrating on such basic pieces of information as student credit hours produced (by various academic levels), student enrollment (head counts), faculty teaching loads, information regarding classroom space, tuition, and the like, CAMPUS is a good example of one way of improving resource allocations in higher education. The CAMPUS system, it should be noted, does not emphasize educational outputs, but rather resource allocation, mainly of a fiscal and physical facilities nature. It is a good example of an MIS designed to improve institutional efficiency, but, at least at the time of this writing, does not appear to be designed to offer college administrators a means of examining their effectiveness.

A good example of a system being designed which hopes to assist institutions (or central agencies)in studying both efficiency and effectiveness is the MIS of the Western Interstate Commission for Higher Education(WICHE) in Boulder, Colorado. The WICHE people are interested not only in the costs of higher education and the best possible means of allocating scarce resources, but also hope to be able to answer the question "what are the outcomes (underlines mine) and products that are produced by those programs and services?"[13] The WICHE rationale is straightforward: "To examine the costs of educational programs with little or no evidence available related to the outputs of those programs offers relatively little advantage to educational decision makers."[14] The WICHE/MIS program is indeed ambitious, for it not only seeks to measure educational outputs and the extent to which higher educational institutions have influenced those outputs, but it goes a step further and wishes to assign dollar signs to the outputs produced. Some of the difficulties in measuring institutional effectiveness are discussed in the following section.

---

[13]Huff, Robert A. Definition and Measurement of the Outcomes and Activities of Higher Education. The Western Interstate Commission for Higher Education, Boulder, Colorado, 1971, p. 1.

[14]Ibid., p.2.

Some of the Many Problems in Measuring Educational Impact

Educational and psychological researchers have been investigating the area of college impact for years, and the methodological problems they have confronted are by now well known to most students of higher education. These include (but are not restricted to) the problem of defining and assessing institutional goals, of relating college effects and college goals, of how (and whether) to develop behavioral objectives for educational institutions, the "lack of variance" phenomenon, and the very difficult problems of inferring causal connections between inputs and outputs in naturalistic settings. Since educational accounting systems attempt to go further and develop ratings of institutional quality on the basis of some of these measures, further problems--particularly nontechnical problems of professional staff morale, interinstitutional competition, and the like--can also be expected to develop, but are beyond the purview of this paper.

The Problem of Defining and Assessing Institutional Goals

Many have been arguing for some time that any evaluation of an institution's effectiveness must take into consideration the institution's goals. The problem, of course, is that too few institutions have really seriously considered what their goals are, and those that have often find that the various members of the college community disagree over what the purposes of the institution should be. It is interesting to note that the recent goals study conducted by Edward Gross and Paul Grambsch used an inventory consisting of 47 goal statements, only 17 of which dealt with "output" goals (teaching students, producing research, providing public service) and the rest dealt with "support" goals, such as academic freedom, involving the faculty in governance of the institution, etc.[15]

---

[15] Gross, Edward W. and Grambsch, Paul V. University Goals and Academic Power. Washington, D. C.: American Council on Education, 1968.

Educational Testing Service has been conducting various studies and literature reviews over the past two years in preparation for the construction of a goals inventory for institutions of higher education. At the time of this writing a preliminary Institutional Goals Inventory (IGI) has been developed and is being "tried out" and modified before being made available for institutional self-study. The preliminary form of the IGI contains 100 statements of plausible institutional goals (e.g., "to help students develop the ability to speak and write effectively," "to strengthen the religious faith of students," "to assist in efforts to achieve and maintain world peace," etc.) to which the respondents--students, faculty, administrators, alumni, trustees, members of the immediate community, or whatever--indicate the extent to which they feel each statement is and should be a goal of the institution. Such an approach makes several things possible. First, while it may be true that divergent groups will never see eye to eye on the major purposes of higher educational institutions, it will at least be possible to quantify the extent of their disagreement and account for it in subsequent studies. Second, the technique provides an interesting measure of discrepancy between what the relevant groups think is and should be highly valued in academia.

However, while instruments such as the one being developed by ETS should be helpful to colleges and universities trying to gain a better perspective of themselves and what they should be doing, the difficulties of trying to assess whether or not they have achieved these goals has just begun.

The Criterion Problem and Behavioral Objectives in Assessing College Impact

Most statements of educational goals--including those in the preliminary IGI described above--are too general in nature to permit precise assessment of whether they have been achieved. How does one determine whether or not the institution has "prepared students for the duties and responsibilities of

citizenship," or "enabled students to develop a set of principles to guide their behavior," or any of a whole series of similar statements which might be found in college catalogues? It was concerns such as these that led to a "movement" toward the development of "behavioral objectives" in education. Behavioral objectives--which are essentially operational definitions--are statements of specific educational objectives in terms of changed student behavior. Such statements lend themselves nicely to direct observation and measurement. (The performance contracting form of educational accounting referred to earlier in this paper relies heavily on behavioral objectives. They contract with school systems not to promote the general level of students' reading ability, but, rather, to improve the class' mean reading score on such and such a test by X number of points.) Behavioral objectives, highly esteemed among educational evaluators for many years, have some serious shortcomings of their own, however. Not least among them stems from their specificity, a characteristic which is at once an advantage and a shortcoming. Because they are highly specific, behavioral objectives permit precise measurement. On the other hand this small precision can be restrictive, in that other highly desirable educational outcomes are omitted. In commenting on this disadvantage of behavioral objectives in the development of mathematics tests, one test specialist has remarked: "...the current statements of behavioral objectives in mathematics for grades K-6 reveal a number of serious defects which would rightly prevent them from being accepted by the mathematics community. The first of these defects seems to result from the energetic attempt to achieve great specificity. The unfortunate consequence of this atomization is that the interrelatedness of mathematical concepts is lost and the statement is a tedious list of very trivial low-level skills...Besides the foregoing, another difficulty in ultimately

stating all the objectives of mathematics instruction behaviorally arises in connection with the desire to develop in students the ability to do original thinking in novel situations. Presumably if these situations and these kinds of thinking were spelled out with the degree of specificity usually found in behavioral objectives, the originality and the novelty would be lost and the objective would 'evaporate in clarity.'" [16]

While the previous criticisms have been directed to behavioral objectives as they relate to mathematics, teachers and testers in other fields are often even less sympathetic to the potential of behavioral objectives. A spokesman for the humanities has chimed in: "This trend (toward the use of behavioral objectives in evaluating school performance) will most likely have disastrous effects on the teaching of English and other subjects in the humanities, for many goals in the humanities either do not naturally result in overt behaviors or result in overt behaviors occurring so far away in time and space from the stimulus presentation that for all practical purposes they are lost to evaluation and will never be counted." [17]

It would be a shame indeed if educational institutions were evaluated in terms of how well their students performed on measures of behavioral objectives which were employed in the first place because they could be measured! Such a situation is much like that of the proverbial tail-wagging-the dog. Cronbach has pointed out that specific behaviors can and should be employed as <u>indicators</u> of constructs (e.g., self-confidence, scientific attitude) but not as the <u>definers</u> of those constructs. Cronbach

---

[16] Myers, Sheldon S. Comments on Behavioral Objectives in Education. Memorandum for the Record, Educational Testing Service, November, 1970.

[17] James Moffett, as quoted in, Myers Miles. The Unholy Marriage--Accountants and Curriculum Makers. <u>American Teacher</u>, 1970, <u>55</u>, 3, November, p. 15.

argues that constructs ought to be the crucial aspect of the evaluation process, where constructs refer to a network of relations or characteristics, but not to specific incidents of behavior. Cronbach goes on to say that "The operationists who want to equate each construct with 'one indicator'...are advocating that we restrict descriptions to statements of tasks performed or behavior exhibited and are rejecting construct interpretations....The writers on curriculum and evaluation who insist that objectives be 'defined in terms of behavior' are taking an ultraoperationalist position, though they have not offered a scholarly philosophical analysis of the issue."[18]

To use as definitions of educational goals--at any level of education--only those criteria which can be measured will almost certainly result in a neat list of narrow and unimportant educational outcomes. To not attempt to state educational objectives in some measurable way tempts educators to rely on the sort of meaningless rhetoric that has characterized college catalogues for many years. The dilemma is a struggle between what Melvin Tumin calls "trivial precision and apparently rich ambiguity,"[19] and it is imperative that institutional administrators and faculty members get with the educational evaluators or "accountants" and attempt to strike a better balance between these two extremes.

That much having been said, it is now just as important to point out that there are probably certain consequences of higher education that will never be measured and perhaps are not measurable. Even after the strict operationalists with their behavioral objectives and the educational

---

[18] Cronbach, Lee J. Validation of Educational Measures, Proceedings of the 1969 Invitational Conference on Testing Problems, ETS, Princeton, N.J., 1969, p. 49.

[19] Tumin, Melvin M. Evaluation of the Effectiveness of Education: Some Problems and Prospects. Interchange, 1, 13 (1970), p. 98.

philosophers with their vague rhetoric agree on objectives which are broader in nature but still measurable, there will remain numerous important educational outcomes which will never be measured in any effective way. Generally, these are the large questions such as "Is higher education really necessary?," "Are the taxpayers getting what they paid for from the publicly supported institutions of higher education?," "Are the educational needs of the state or region being satisfied?," and so on. None of these questions, at least as they are phrased here, can be answered by the most sophisticated evaluation or educational accounting. At least not until each of these "large" questions are split into a great many more "specific" questions. This process of "clarification," however, according to Tumin again, very often proves "to be one of selecting a very few of the many constituent facets of those questions and focussing on those alone, hoping that those fragments will somehow 'represent' or 'stand for' the large whole, such as is implied in 'serving the needs' or 'preparing the children,' or other comparable 'holistic' phrases. In short, if reliable measurements are to be demanded, it is indispensable that the 'whole' impact in which we are always interested, be broken up into fragments, and certain selected aspects of those 'whole' taken under study, while the many other fragments and the 'wholeness' are once again put aside."[20]

This should not be interpreted to mean that educational evaluators should despair of developing useful, reliable, comprehensive measures of educational outcomes. Many have already been developed and efforts to develop better ones should continue. But those who work on such problems should be guided by the realistic awareness that the "large" questions regarding American higher education will probably not be answered through their efforts.

---

[20] Ibid., p. 98.

The Lack of Variance Problem and the Need for Multiple Criterion Measures

Almost all proponents of educational accountability tend to favor a "value added" concept. That is, institutions should be judged not by their outputs alone, but by their outputs relative to their "inputs." The students' final "standing" with regard to various characteristics would not be as important as their changes (usually gains) during the college years. A rather typical point of view is the following: "What has the student attained in relation to his capability at the starting point? This concept approximates educational value-added....According to this view, an educational process which moved the student from the lowest quartile of high-school achievement to the second quartile of college-graduate achievement would be accomplishing something tremendous, whereas the college which accepted students only from the top decile of high school achievement and delivered them into the top decile of college achievement would be doing relatively much less."[21]

Such a view--and again, let me emphasize that it is a view widely held--makes the assumption that educational institutions are potentially very powerful change agents, capable of having a great deal of impact on both the cognitive and non-cognitive attributes of all who pass through their doors. It is further assumed that colleges differ widely in the amount of impact they have. The accuracy of such a view, however, is highly questionable. Indeed, most of the evidence suggests that it is downright naive, for educational institutions at all levels appear to differ very little in terms of the amount of impact they have on their students after controls are made for general mental ability, socioeconomic status (SES), and other important background factors outside the purview of the

---

[21] Balderston, Frederick E. Thinking About the Outputs of Higher Education, The Outputs of Higher Education: Their Identification, Measurement, and Evaluation, Western Interstate Commission for Higher Education, Boulder, Colo., July, 1970, (p. 14).

formal educational institution. For example, numerous proponents of the "value-added" concept in educational accountability argue that one good criterion for institutional quality would be their students' standing on standardized tests of educational "attainment," after controls have been made for educational aptitude at the time of entry into college. Very often specific suggestions are made for use of one of the national college admissions tests (the SAT of the College Entrance Examination Board or the ACT) as the input measure and scores on one of the Area Tests of the Graduate Record Examinations (GRE) as the output measure.[22] At first blush, such an approach seems quite sensible. The problem, however, is that the correlation between college means on these measures is so high (often in the .90's) that there is generally very little variance left that the schools can influence. Obviously, the overlap between the input and output measure varies somewhat depending on the specific measures chosen for the study, but it is true that any two measures of academic aptitude or achievement (and the distinction between the two is often very fuzzy indeed!) will correlate quite highly. This is generally referred to as the "g" factor by psychologists, reflecting the general nature of cognitive skills required on such tests. While there is some variance remaining (that is, some test performance that cannot be attributed to this general factor), this portion of the variance can usually be best explained by differences in SES. Only a tiny portion of differences in cognitive tests scores remains that cannot be explained by one of these two factors. Assuming that the balance is all caused by differences in educational experiences (an unlikely assumption), the point is that there is precious little opportunity for educational influences to be regarded as very important in explaining differences in

---

[22]Technically, the Graduate Record Examinations now refer exclusively to the aptitude and achievement measures (Advanced Tests) used for graduate school admission. The tests formerly known as the GRE Area Tests are now part of ETS' new undergraduate Program for Counseling and Evaluations (UP).

student performance on such measures. This is not meant to suggest that formal education has no influence on its students. Notice that the comparison is always between schools and seldom (if ever) based on a school vs. no-school dichotomy. Schools may have some influence, but the degree of their influence is almost indistinguishable from each other. This seems to be true not only in the area of cognitive traits, but for various non-cognitive (e.g., attitudes and values) traits as well. Researchers have been interested in the question of college impacts on students' attitudes and values for years, and have usually come to the conclusion that while students definitely change during the college years, it is extremely difficult to associate those changes with colleges possessing certain characteristics. In the most comprehensive summary of college impact research that has ever been published, Feldman and Newcomb point out that "the degree and nature of different colleges' impacts vary with their student inputs," and later, "In the absence of more complete data, we offer it only as a likely hypothesis that those characteristics in which freshman-to-senior change is distinctive for a given college will also have been distinctive for its entering freshmen...(their underlines)"[23]

Part of the difficulty in discovering differential cognitive impact of educational institutions may be attributable to a lock-step methodology which is clouding real impact differences. Given the nature of most tests of cognitive attributes used in such research, it probably shouldn't be too surprising that they do not turn up large educational differences. These tests are almost always constructed so as to be widely appropriate and sufficiently general in nature to ensure their appropriateness for many educational experiences. Yet herein lies part of the evaluative problem. Criterion measures designed to be

---

[23]Feldman, Kenneth A., and Newcomb, Theodore M. The Impact of College on Students. San Francisco: Jossey-Bass Inc., 1969, pp. 327 and 328.

broadly applicable may well be too general in nature to measure the specific outcomes of educational experiences at a local level. Educational evaluators may have to turn, instead, to achievement examinations geared especially to syllabi used in specific college courses if they are to turn up indices of college effects. Such a procedure makes it difficult, however, to conduct inter-institutional comparisons, often felt to be the central and most important feature of educational accounting systems. Thus, we are back to the problems suggested earlier; measures of a general nature yield little or no inter-institutional variation, while measures geared to the program of a specific department or institution do not allow for multi-college comparisons. Yet, the inter-institutional comparisons are useless if they fail to reveal meaningful differences, and so the specifically-designed criterion measures may be the only reasonable solution.

Reliance on a far greater variety of criterion measures (outcomes measures) would also seem to be desirable. This is particularly true during a period of what seems to border on universal higher education. With students of varying backgrounds, skills, interests, and objectives attending our institutions of higher education, it seems imperative that we begin to examine criteria other than some form of "intellectuality," which, like it or not, can no longer be regarded as the primary purpose of most higher educational institutions.

As with other aspects of the educational evaluation paradigm, however, it is easy to talk about the need for a variety of criterion measures and much harder to come up with them. Social conscience, heightened awareness, various kinds of "appreciation," attitudes and values, citizenship, moral sensitivity--all these and more have been mentioned as projected outcomes of certain colleges. Measures of these variables will surely not be a simple task, but there is some

reason for optimism. As long as it is remembered that such measures would serve as indicators (and not definers) of desired educational constructs, the development of the inventories and materials would be a difficult, time-consuming, expensive--but definitely possible and hopefully worthwhile task.

The Problem of Inferring Effects in Naturalistic Settings

In order to use output measures of student performance to compare the effectiveness of educational programs, adjustments must be made for preexisting differences among the groups. These adjustments are the crux of the "value added" concept discussed at the beginning of the preceding section. Unfortunately, there is no guarantee that any of the frequently used means of making adjustments such as matching, using difference scores, analysis of covariance, or other regression techniques will result in an appropriate adjustment. As stated by Lord, "...there simply is no logical or statistical procedure that can be counted on to make proper allowances for uncontrolled preexisting differences between groups."[24]

There are two major aspects to the problem of making adjustments, (1) the identification of all of the relevant variables for which adjustments are needed, and, (2) the estimation of the magnitude of the adjustment that should be made for the variables once they are identified. It seems clear that allowances should be made for differences in student aptitudes at time of entrance into the program. Certain background characteristics such as SES are also natural candidates. However, there are many other potentially important differences among entering students that are typically ignored or not thought of (e.g., motivation, sex, age). Adjustments also are needed for institutional characteristics that can not be controlled by the institution.

---

[24]Lord, Frederic M., A paradox in the interpretation of group comparisons, Psychological Bulletin, 1967, 68, 21-38 (p. 305).

Given a set of variables for which adjustments are desired, there remain several sources of error that can result in biased adjustments. Specification errors and errors of measurement can both bias the comparisons of preexisting groups. The failure to include a variable in the model that is related either to the output or other control variables and on which there are preexisting differences among groups would be a specification error that would result in bias. Similarily, unreliability in the control variables will result in biased adjustments when the groups differ on these variables initially. As Astin points out, the most likely result of these shortcomings is to misleadingly indicate college effects when, in fact, there may be none.[25]

Conclusions

These problems suggest that evaluating differential college impact may not be possible at all, or, at best, that it will be some time before it can be done very well. The real difficulty is not so much in developing new, reliable, relevant criterion measures. That will be difficult, of course, but certainly no insurmountable task. The problem will be in demonstrating differential college effects on these various criteria. Obviously, criteria which do not yield meaningful between-college differences in institutional effects will not be useful for evaluating the effectiveness of those institutions.

---

[25]Astin, Alexander W. The Methodology of Research on College Impact, (Part One) Sociology of Education, 1970, 43, 223-253.

For this reason, it might make sense to begin at the beginning and help institutions do better in the area of institutional efficiency. Immediate attention to the development of management information systems which would permit college administrators to base everyday administrative decisions on continually updated facts about the institution would be welcome service, and one which could be done rather soon. Forecasting detailed space requirements, calculating the number of faculty members needed for different enrollments, showing how operating costs would increase or decrease with a change in certain class scheduling techniques, consideration of alternative staffing policies such as teaching loads, tenure, and the like--all these very important aspects of institutional functioning could be based on facts routinely gathered and summarized, if only more institutions knew how to do it. MIS specialists could do higher education a great service in this educational efficiency area.

While that is being done, other specialists could continue to grapple with the problems of assessing the outcomes of higher education. It would indeed be unfortunate to turn all our attention to the area of educational efficiency, and ignore the question of college impact, thus taking part in what Selznick calls the "cult of efficiency" which overstresses means and totally neglects ends.[26] But the question is whether, given the limitations outlined earlier, it makes sense to hold institutions "accountable" for their effectiveness just yet, and whether the efficiency of operations couldn't be vastly improved while the effectiveness question is being considered.

---

[26]Selznick, P. Leadership in Administration. New York: Harper & Row, 1957.

In any event, whether dealing with operational efficiency or educational effectiveness, it would be well to remember that education is a social process and will inevitably resist simplistic evaluations of its results. As Henry Dyer has said:

> The term educational accountability, as used most recently by certain economists, systems analysts, and the like, has frequently been based on a conceptualization that tends, by analogy, to equate the educational process with the type of engineering process that applies to industrial production....It must be constantly kept in mind that the educational process is not on all fours with an industrial process; it is a social process in which human beings are continually interacting with other human beings in ways that are imperfectly measurable or predictable. Education does not deal with inert raw materials, but with living minds that are instinctively concerned first with preserving their own integrity and second with reaching a meaningful accommodation with the world around them. The output of the educational process is never a "finished product" whose characteristics can be rigorously specified in advance; it is an individual who is sufficiently aware of his own incompleteness to make him want to keep on growing and learning and trying to solve the riddle of his own existence in a world that neither he nor anyone else can fully understand or predict.[27]

Dyer's quote, perhaps more than all the limitations discussed earlier in this paper, serves to emphasize that the problems involved in assessing institutional effectiveness and developing objective criteria for accountability will continue to be hard problems. They are precisely the problems, however, that must be tackled with the best people and the best methods available if higher education is going to serve us well.

---

[27]Dyer, Henry S. Toward Objective Criteria of Professional Accountability in the School of New York City, Phi Delta Kappan, 1970, LII, 4, December, p. 211.